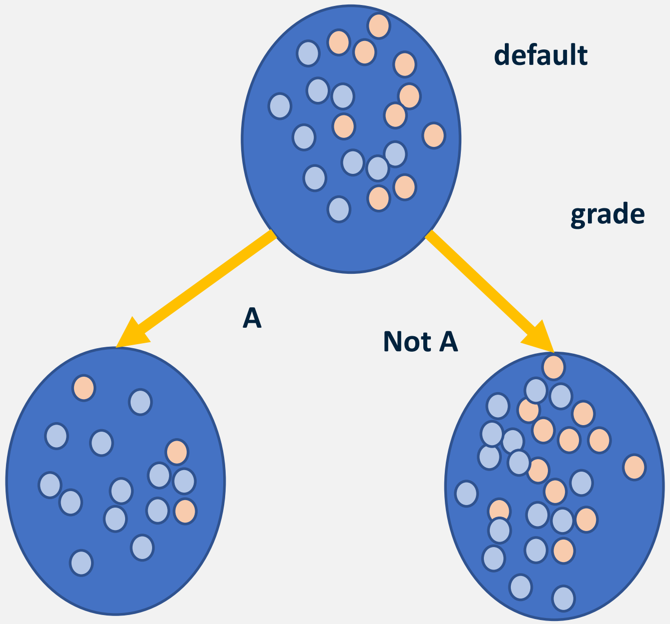
LendingClub

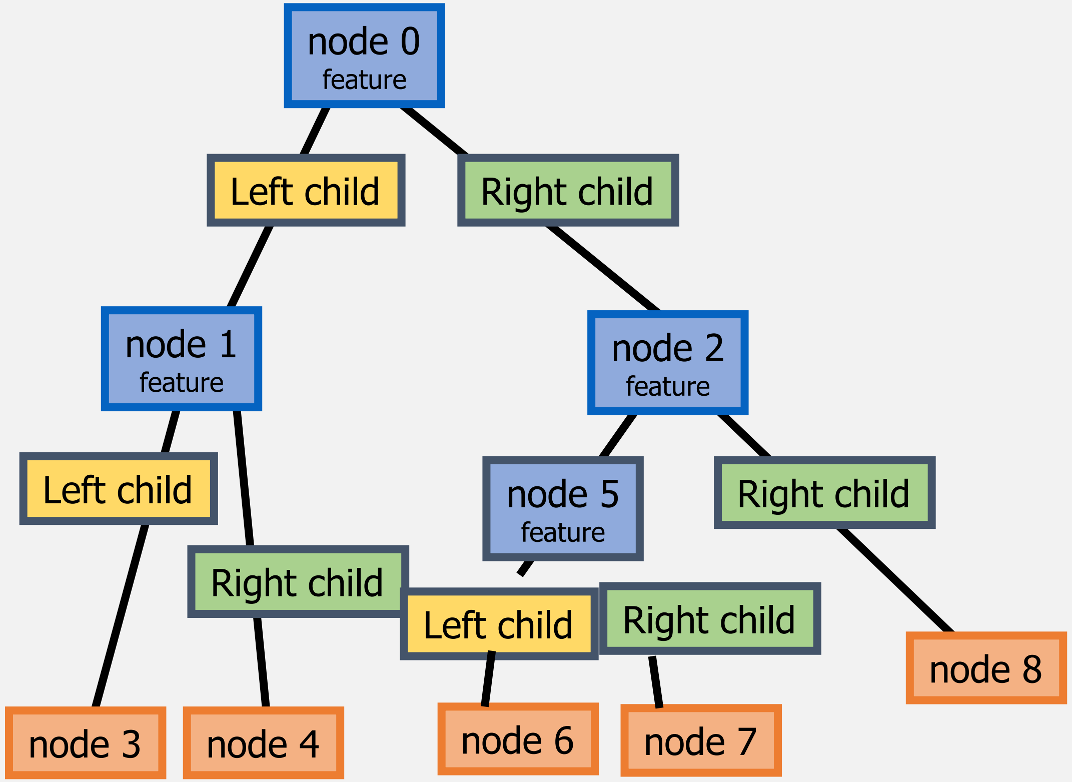
Fitting a Model

Lecture plan

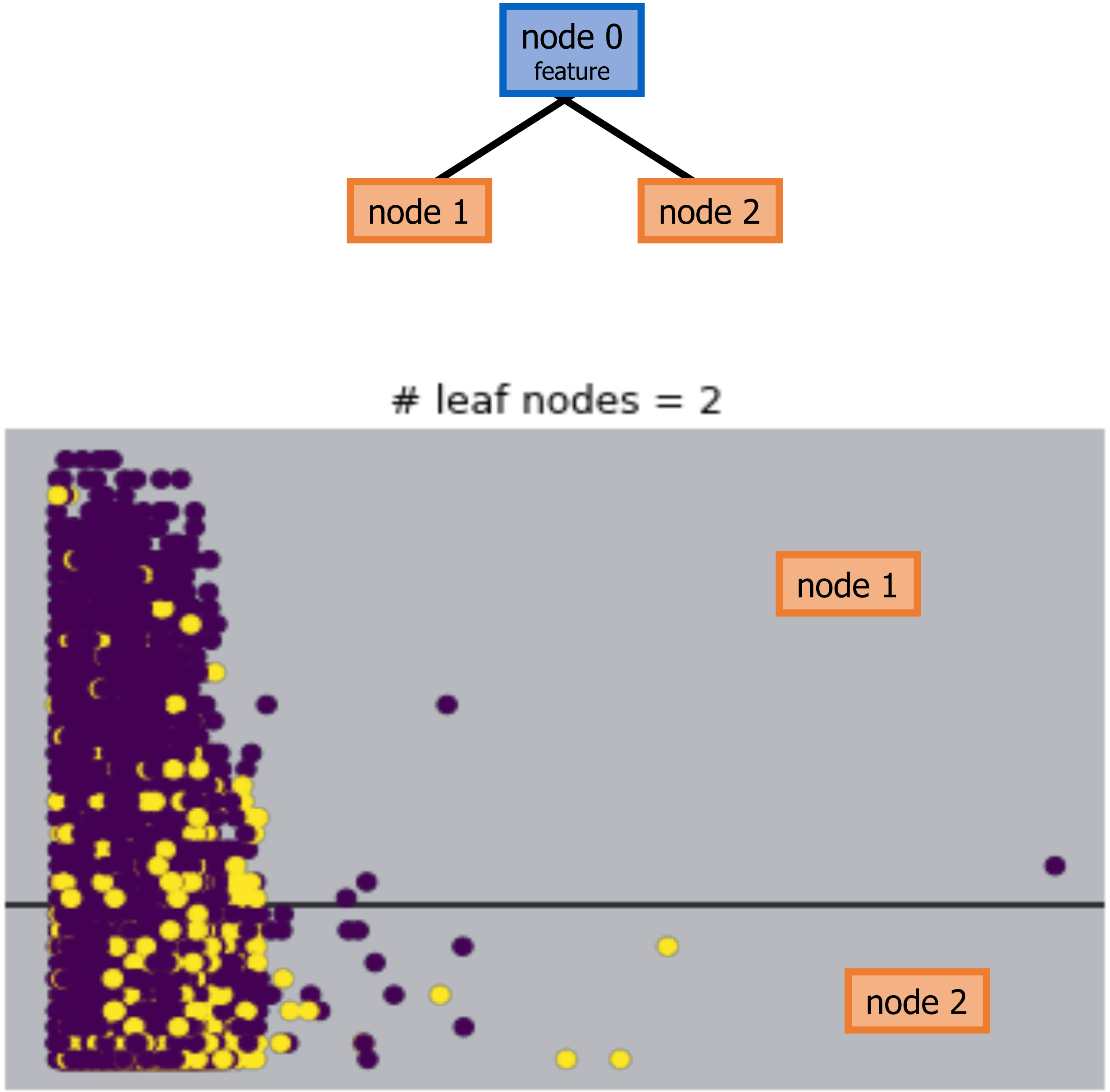
- Decision Tree Geometric Interpretation
- Labeling the leaves
- Loss Function
- Regression Tree
- Linear Regression

Decision Tree
Where we left off

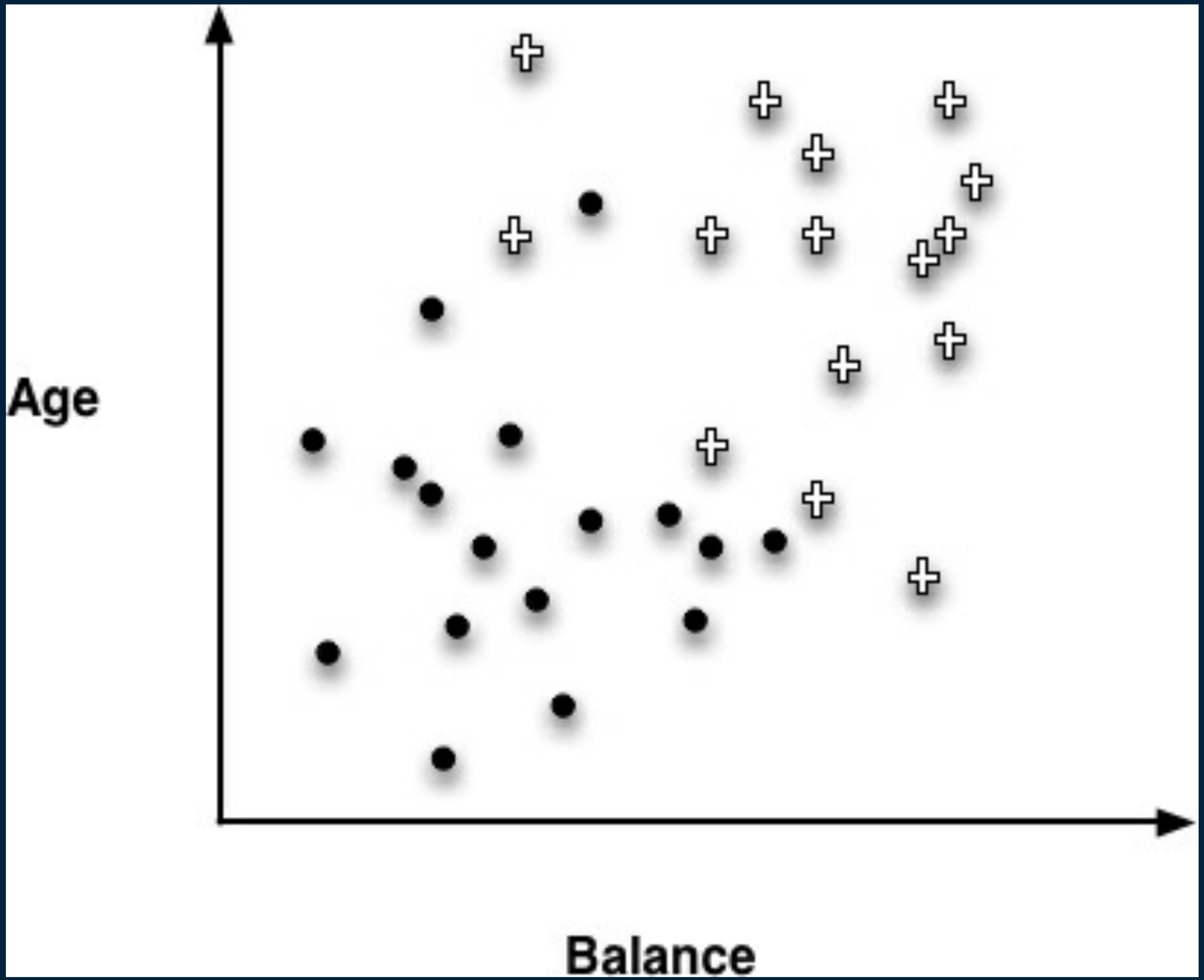




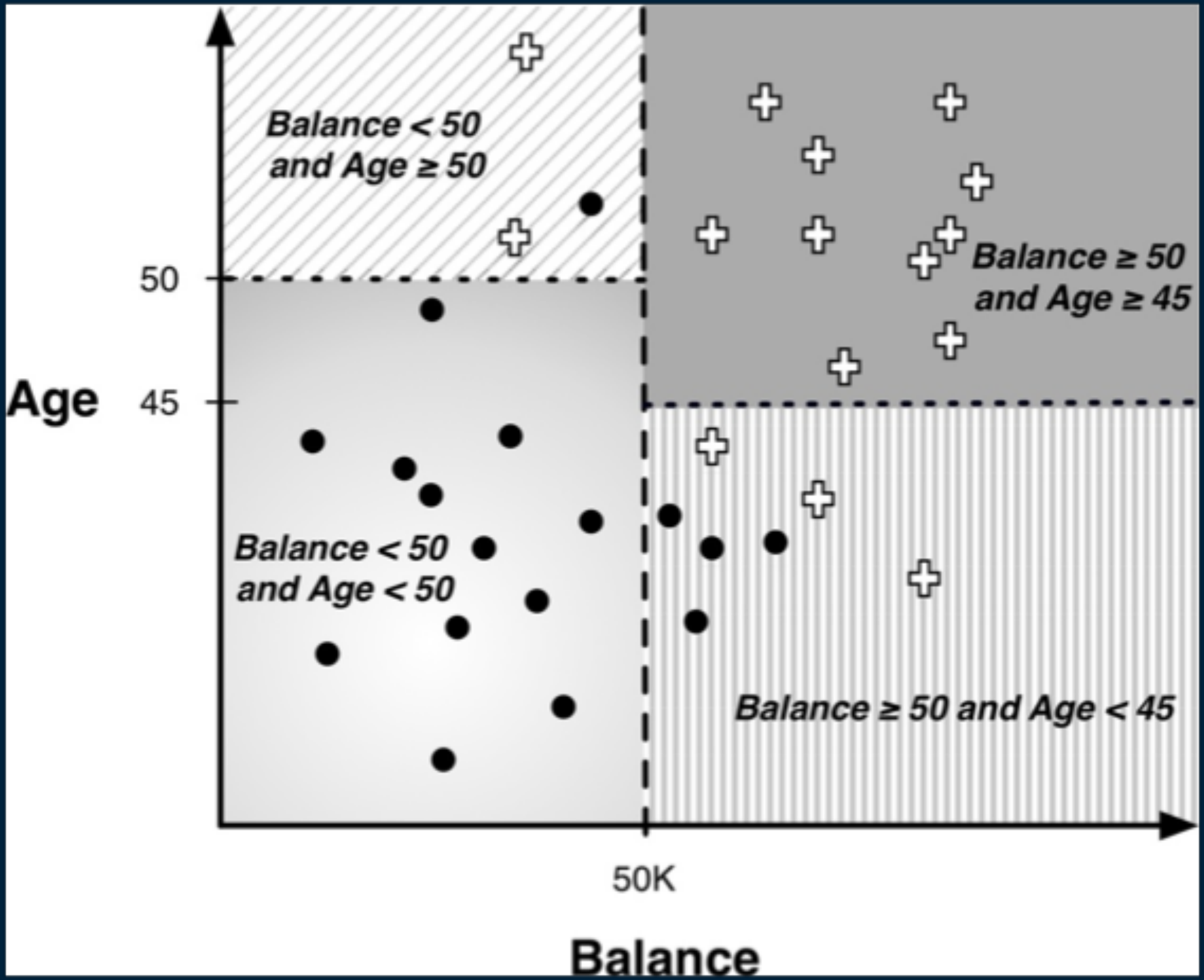
Geometric Interpretation



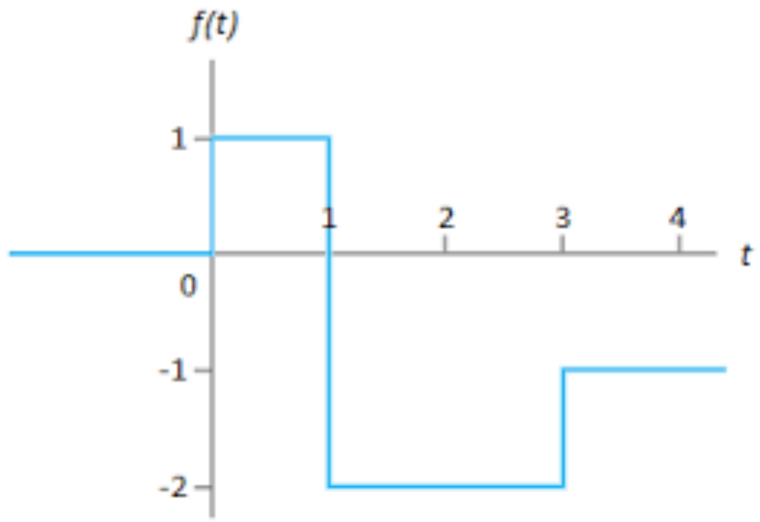
Geometric Interpretation

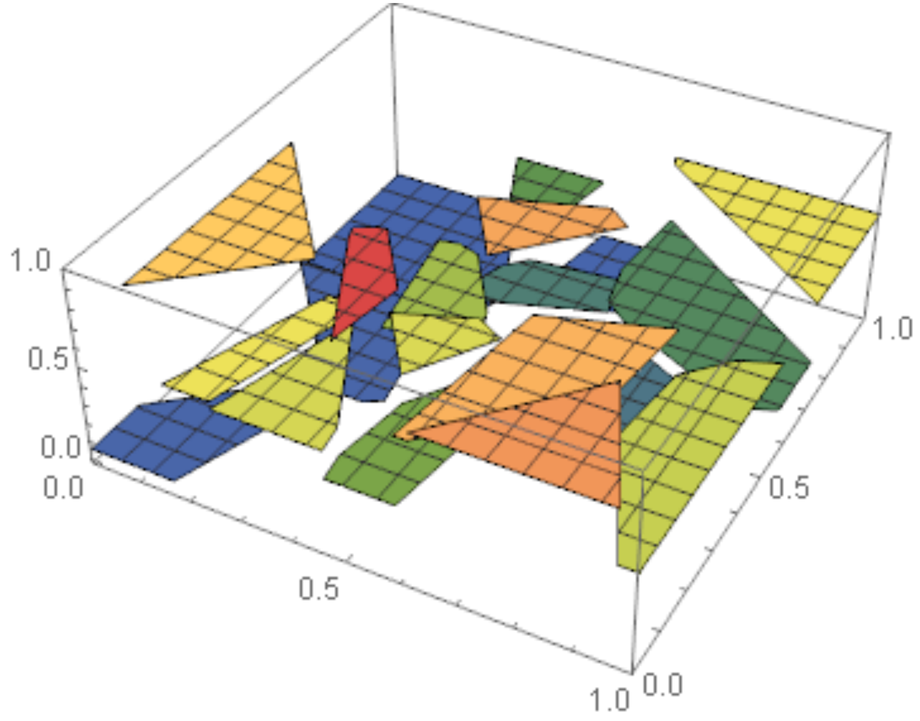


Geometric Interpretation



Geometric Interpretation





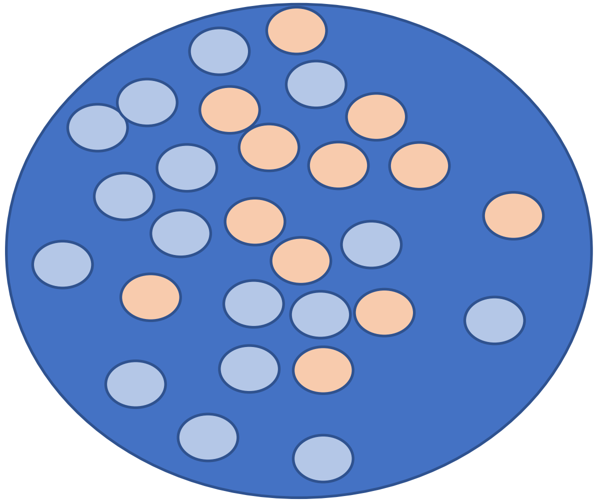
Let $(x_1, \dots, x_k)$ a tuple in the feature space, the a decision tree defines a function

$$y = Tree(x_1, \dots, x_k)$$

- The decision tree induces a uni/multivariate piecewise constant function , aka a bar function

Geometric Interpretation

- The features of a particular individual $(x_1, \dots, x_k)$ associates the individual to a group of individuals with features in a subspace in the feature space
- We must now associate with this group a number or label that will be the forecast for this group

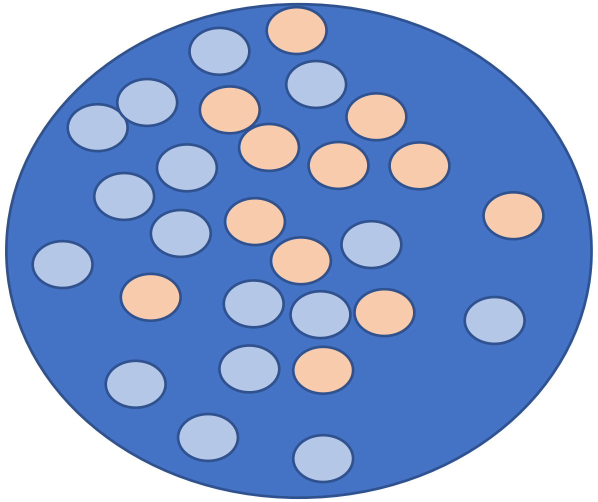


In the setting of an urn with red and blue balls

- Our objective is to forecast a random draw from the urn

What would be a good forecast for the outcome?

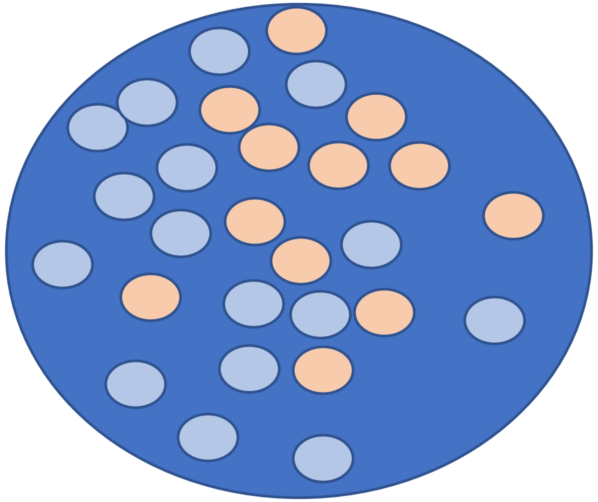
Stat 101



What would be a good forecast depends on what is our ***objective***

- **[Case 1]**: Guessing game where players are rewarded $1 iff their forecast is identical to the draw
  - Objective:
    - maximize payoff / minimize loss
  - 0-1 win/loss
- **[Case 2]**: Guessing game with admission where players are rewarded $1 less admission iff their forecast is identical to the draw
  - Objective:
    - play/not play
    - maximize payoff / minimize loss
  - Continuous win/loss

0-1 loss



[Case 1]: Guessing game where players are rewarded \$1 iff their forecast is identical to the draw

- Objective – maximize the payoff
- 0-1 win/loss

Let $\widehat{\boldsymbol{x}}_{\mathbf{1}} \dots \widehat{\boldsymbol{x}}_{\boldsymbol{k}}$ be a sample of individuals in the feature space, $\widehat{\boldsymbol{y}}_{\mathbf{1}} \dots \widehat{\boldsymbol{y}}_{\boldsymbol{k}}$ be the corresponding labels of these individuals and l the forecast for a draw from the urn, we would like to optimize the following objective function

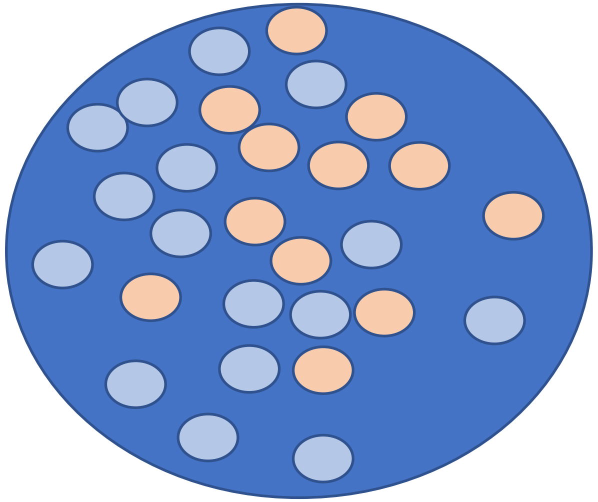
$$L(\hat{y}) = \sum_i \chi_{\widehat{y_i} \neq l}$$

Optimizing Forecast: $L(\hat{y}) = Maj(\widehat{\boldsymbol{y}}_{\mathbf{1}} \dots \widehat{\boldsymbol{y}}_{\boldsymbol{k}})$

L_2 - loss

[Case 2]: Guessing game with admission where players are rewarded $1 less admission iff their forecast is identical to the draw

- Objective:
  - play/not play
  - maximize payoff / minimize loss
- Continuous win/loss



Let $\hat{\boldsymbol{y}}_1 \ldots \hat{\boldsymbol{y}}_k$ be as before and p the charge for playing the game

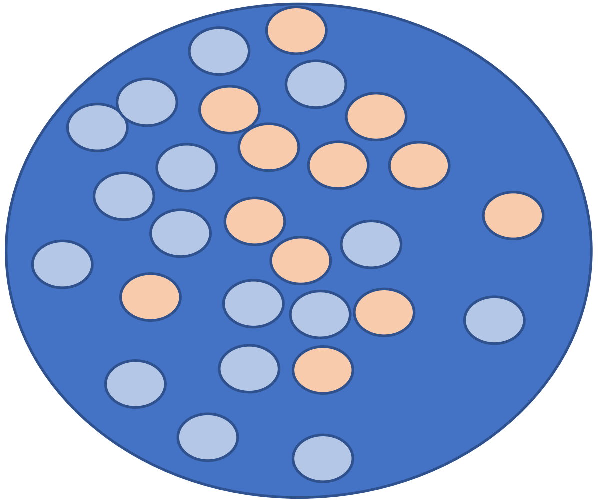
$$L(\hat{y}) = \sum_i (\hat{y}_i - p)^2$$

Optimizing Forecast: $L(\hat{y}) = mean(\hat{\boldsymbol{y}}_1 \ldots \hat{\boldsymbol{y}}_k) = \frac{1}{k} \sum_i \hat{\boldsymbol{y}}_k$

L_1 - loss

[Case 2]: Guessing game with admission where players are rewarded $1 less admission iff their forecast is identical to the draw

- Objective:
  - play/not play
  - maximize payoff / minimize loss
- Continuous win/loss



Let $\hat{\boldsymbol{y}}_{\mathbf{1}} \ldots \hat{\boldsymbol{y}}_{\boldsymbol{k}}$ and p be as before

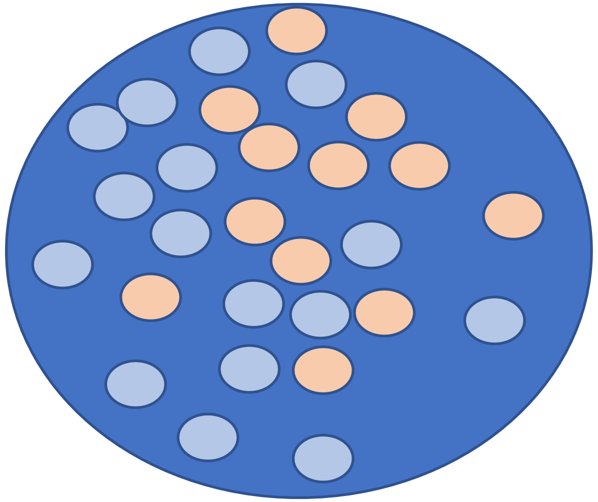
$$L(\hat{y}) = \sum_i |\hat{y}_i - p|$$

Optimizing Forecast: $L(\hat{y}) = median(\hat{\boldsymbol{y}}_{\mathbf{1}} \ldots \hat{\boldsymbol{y}}_{\boldsymbol{k}})$

Maximal Likelihood

[Case 2]: Guessing game with admission where players are rewarded $1 less admission iff their forecast is identical to the draw

- Objective:
  - play/not play
  - maximize payoff / minimize loss
- Continuous win/loss



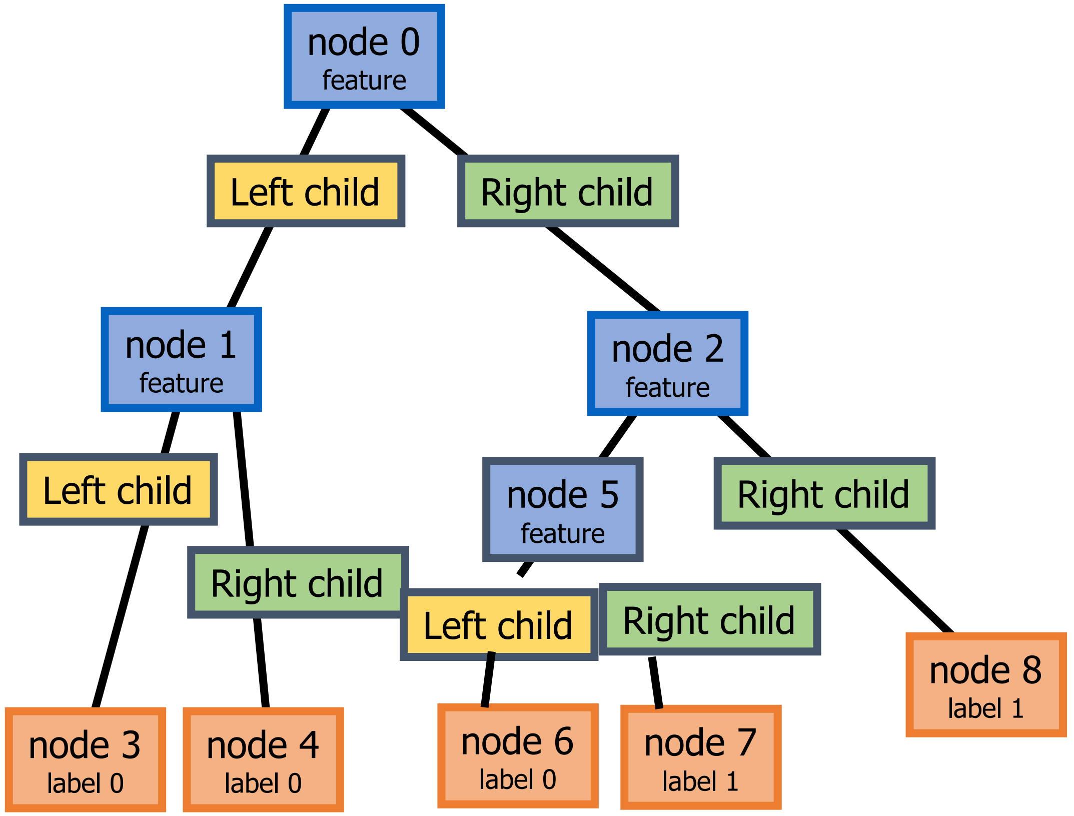
Let $\widehat{\boldsymbol{y}}_{\mathbf{1}} \ldots \widehat{\boldsymbol{y}}_{\boldsymbol{k}}$ and p be as before and $\mathrm{k} = \sum_i \widehat{\boldsymbol{y}}_i$

$$L(\hat{y}) = \mathrm{m} \cdot \log(p) + (k - m) \cdot \log(1 - p)$$

Optimizing Forecast: p=$\frac{m}{k}$=median($\widehat{\boldsymbol{y}}_{\mathbf{1}} \ldots \widehat{\boldsymbol{y}}_{\boldsymbol{k}}$)

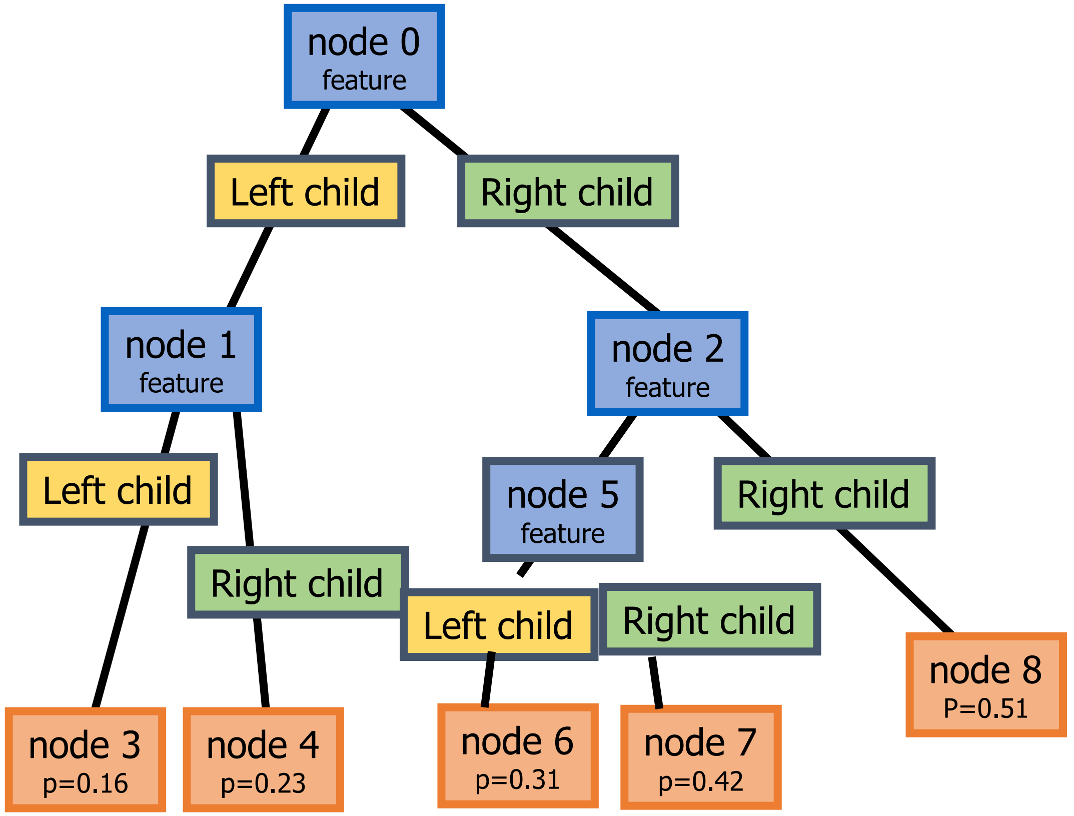
Decision Tree

- Each leaf labeled 0/1 based on optimizing 0-1 loss on subspace of feature space

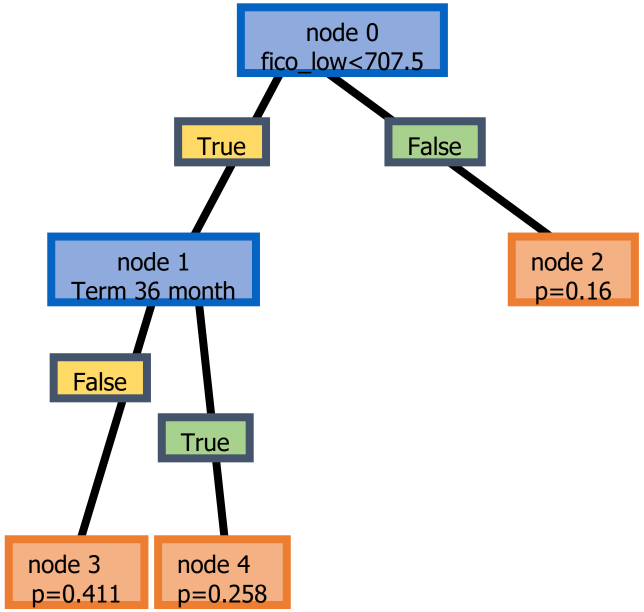


Regression Tree

- Each leaf given the maximal likelihood probability on subspace of feature space
- ID3 algorithm optimizes on maximal likelihood loss



Regression Tree



The binary tree structure has 5 nodes with thefollowing structure:

node=0 test: if fico_range_low<=707.50 goto node 1 else to node 2

node=1 test: if term:: 36 months<=0.50 goto node 3 else to node 4

node=2 leaf

node=3 leaf

node=4 leaf

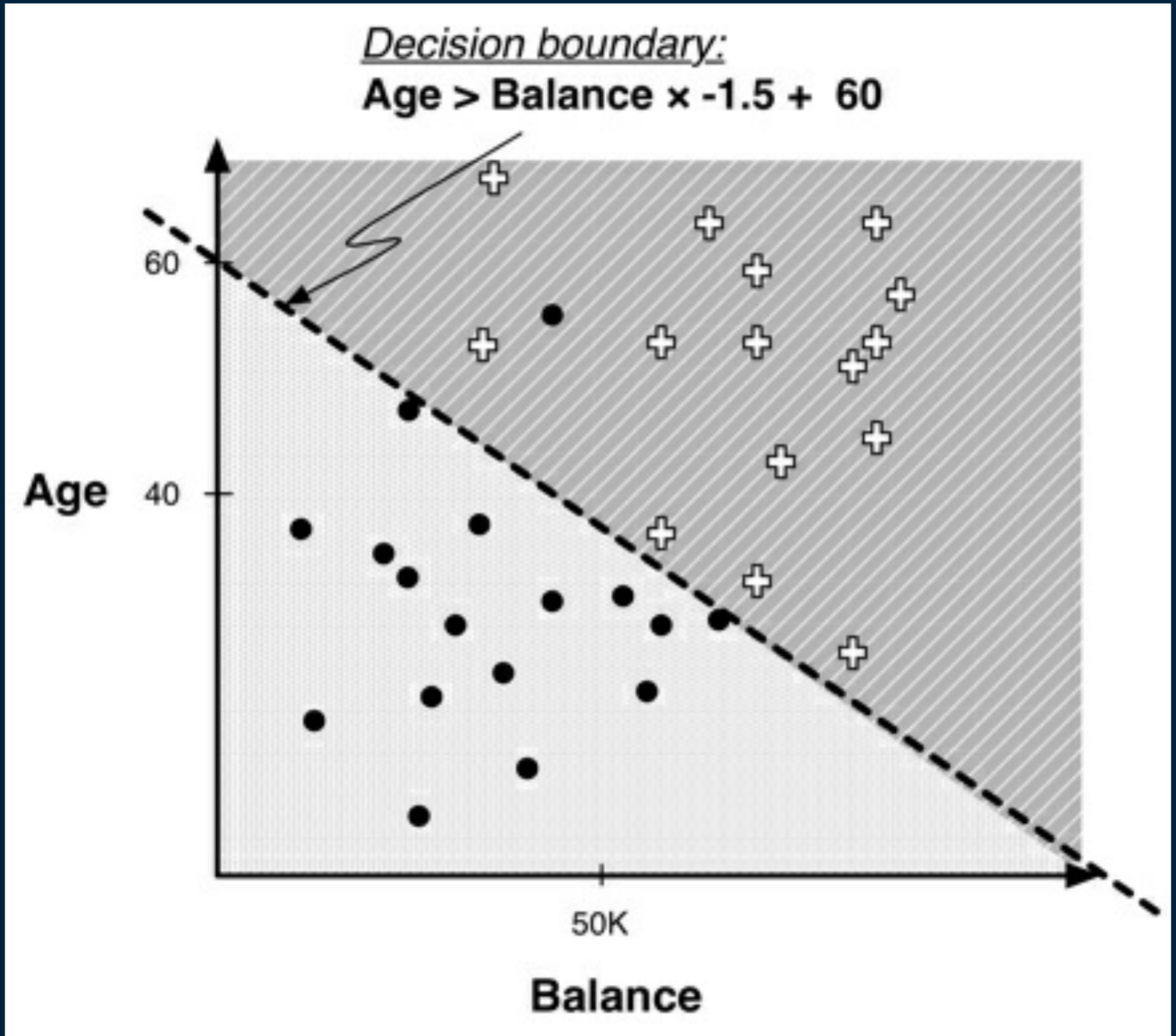
In [235]:
```
leave_id = decision_tree.apply(X_test)
pd.DataFrame({'leave':leave_id,'label':y_test}).groupby(['leave']).mean()
```

Out[235]:

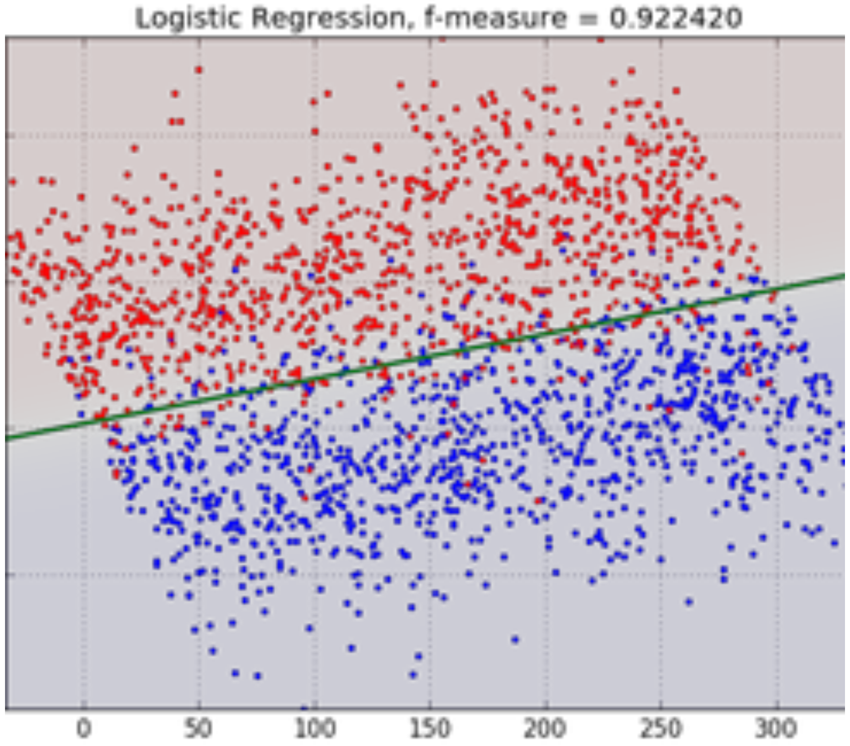
leave	label
2	0.162920
3	0.411017
4	0.258824

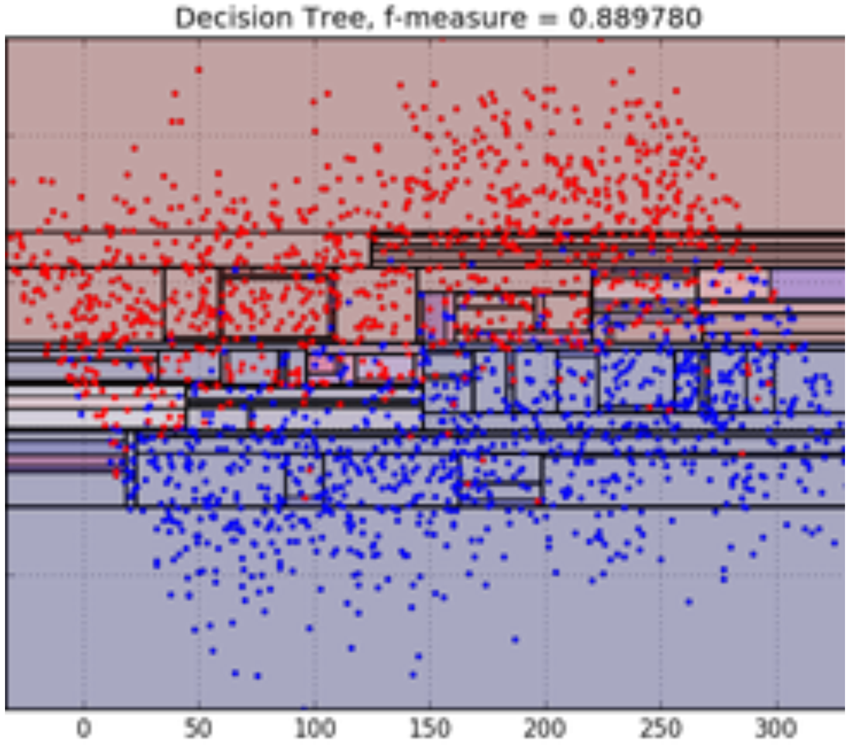
Geometric Interpretation

Linear Regression



Geometric Interpretation





Geometric Interpretation